Systematic Exploration of S2D Predictability Limits

Steve Yeager
and colleagues

NSF NCAR
Boulder, CO, USA

Predictability Limits

- Observations: $$X = \chi + x = \chi_f + \chi_i + x$$

Potentially Predictable

Noise

➔ real world predictability: $$p = {\sigma_\chi^2}/{\sigma_X^2}$$

➔ inherent prediction skill limit: $$r_{max} = \sqrt{p} \leq 1$$

Theoretical, not quantifiable!

Predictability Limits

- Observations: $\quad X = \chi + x = \chi_f + \chi_i + x$
- Initialized Predictions: $\quad Y = \psi + y = \psi_f + \psi_i + y$

Potentially Predictable
Noise

Large ensemble limit $(m \to \infty)$

➔ ensemble average: $\quad \sigma_{Y_a}^2 = \sigma_{\psi_f}^2 + \sigma_{\psi_i}^2 + \frac{1}{m}\sigma_y^2 \to \sigma_{\psi_f}^2 + \sigma_{\psi_i}^2 = \sigma_\psi^2$

➔ model potential predictability: $\quad q = \sigma_{Y_a}^2 / \sigma_Y^2$

Quantifiable, but unclear
how it relates to real world

➔ model skill at predicting real world: $\quad r \le r_{max}$

Validation against observations
is only reliable approach for
estimating r_{max}

Seasonal Hydroclimate Prediction

- CESM2-SMYLE system (~10K sim-years)

- 1979-2021 verification against GPCPv2.3

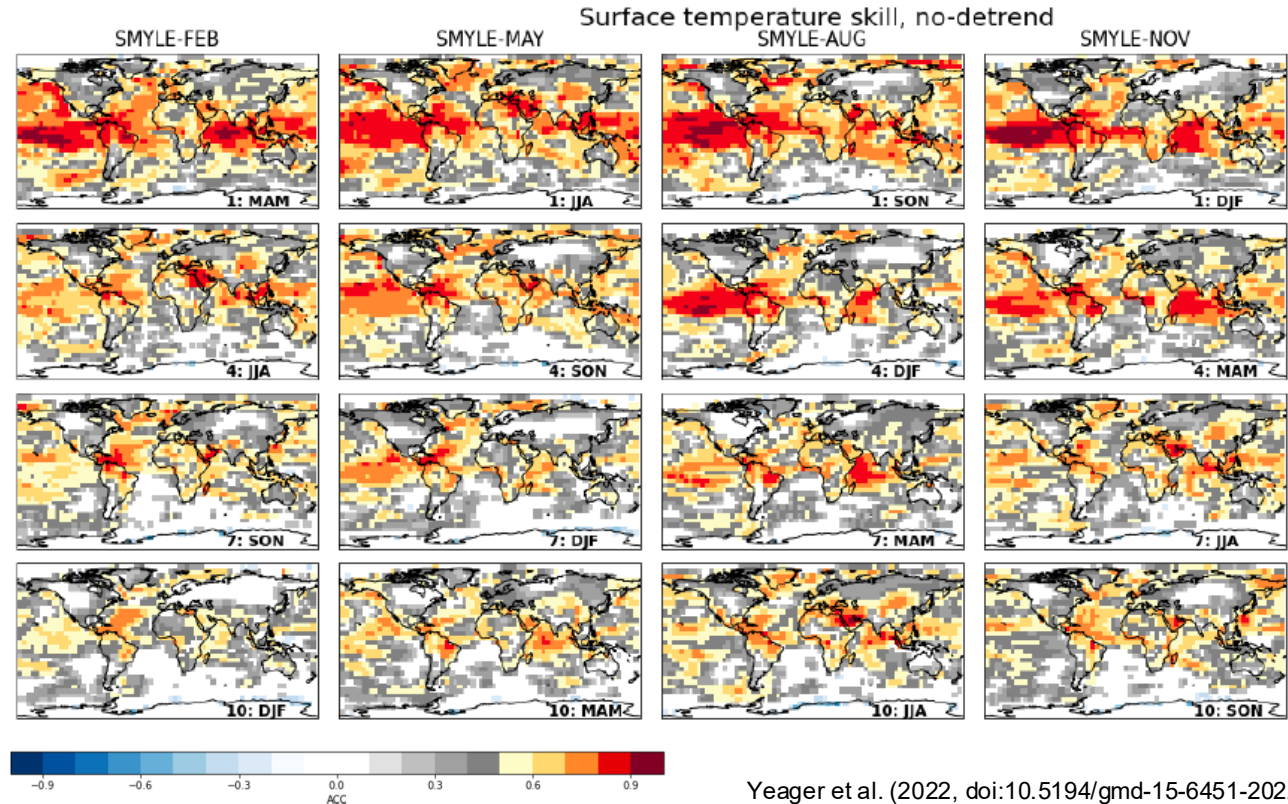- Low skill over CONUS even at short leads

Precipitation skill, no-detrend

SMYLE-FEB   SMYLE-MAY   SMYLE-AUG   SMYLE-NOV
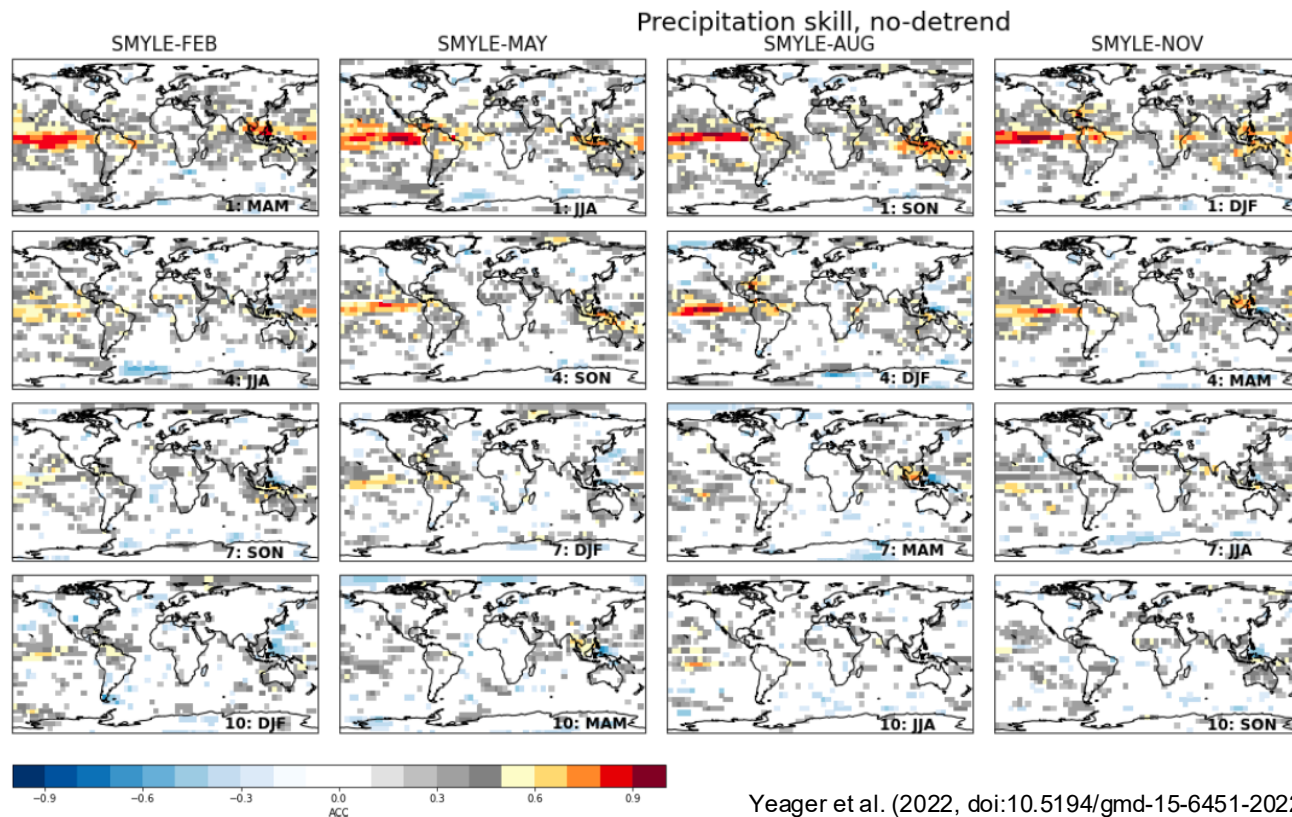
Yeager et al. (2022, doi:10.5194/gmd-15-6451-2022)

Seasonal Hydroclimate Prediction
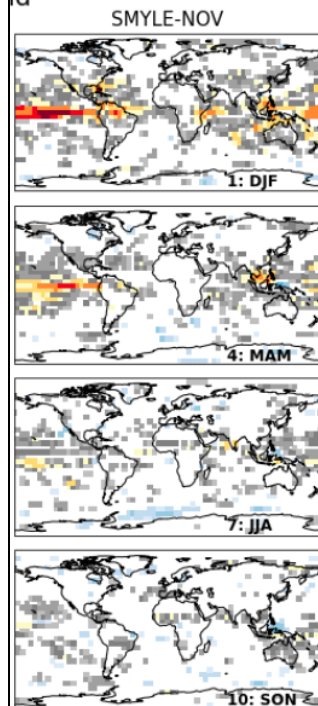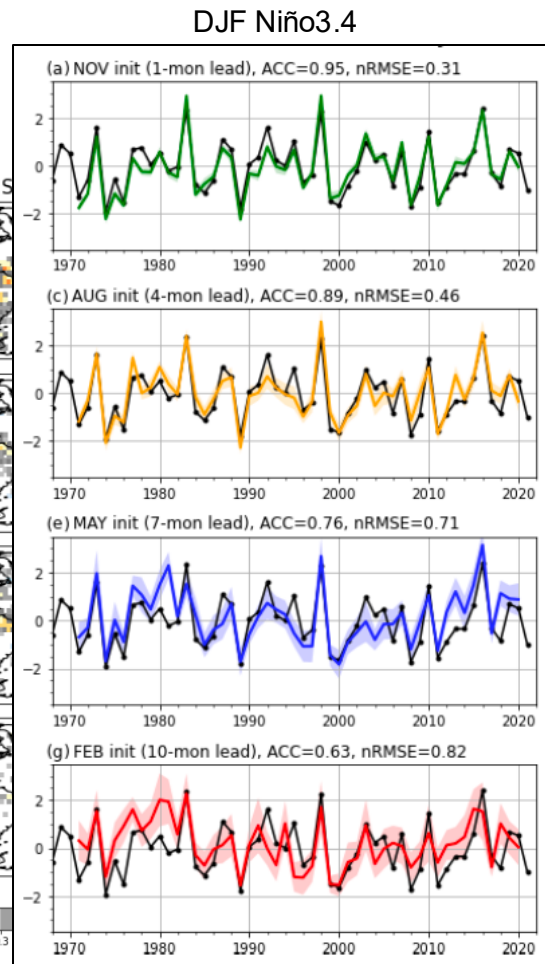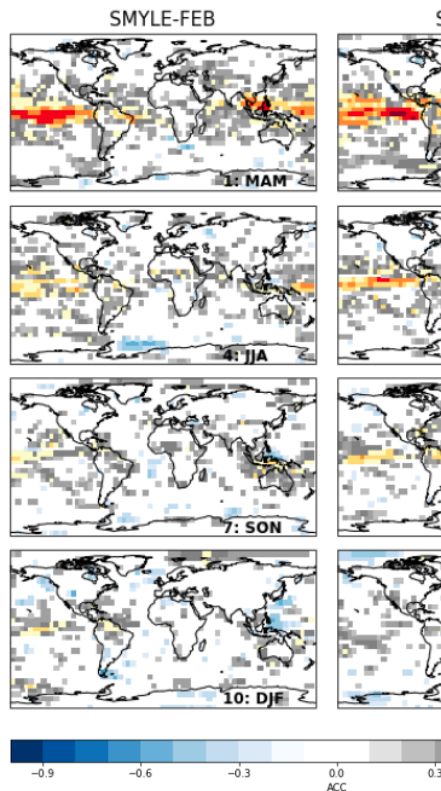
- 1970-2020 verification against (CRU-TS4.05, HadISST)

Surface temperature skill, no-detrend

Seasonal Hydroclimate Prediction

- CESM2-SMYLE system (~10K sim-years)

- 1979-2021 verification against GPCPv2.3

- Low skill over CONUS even at short leads

**Are we underestimating real world hydroclimate predictability? Why and by how much?**

DJF Niño3.4

(a) NOV init (1-mon lead), ACC=0.95, nRMSE=0.31

(c) AUG init (4-mon lead), ACC=0.89, nRMSE=0.46

(e) MAY init (7-mon lead), ACC=0.76, nRMSE=0.71

(g) FEB init (10-mon lead), ACC=0.63, nRMSE=0.82

SMYLE-FEB

SMYLE-NOV

1: MAM

4: JJA

7: SON

10: DJF

1: DJF

4: MAM

7: JJA

10: SON

ACC

doi:10.5194/gmd-15-6451-2022)

Model/System Sensitivity

- E3SMv2.1-SMYLE system (~10K sim-years)

- Different component models but uses similar methods as CESM2-SMYLE

- Early analysis suggests no systematic, significant* skill difference for Niño3.4 or hydroclimate over CONUS

  Nov init
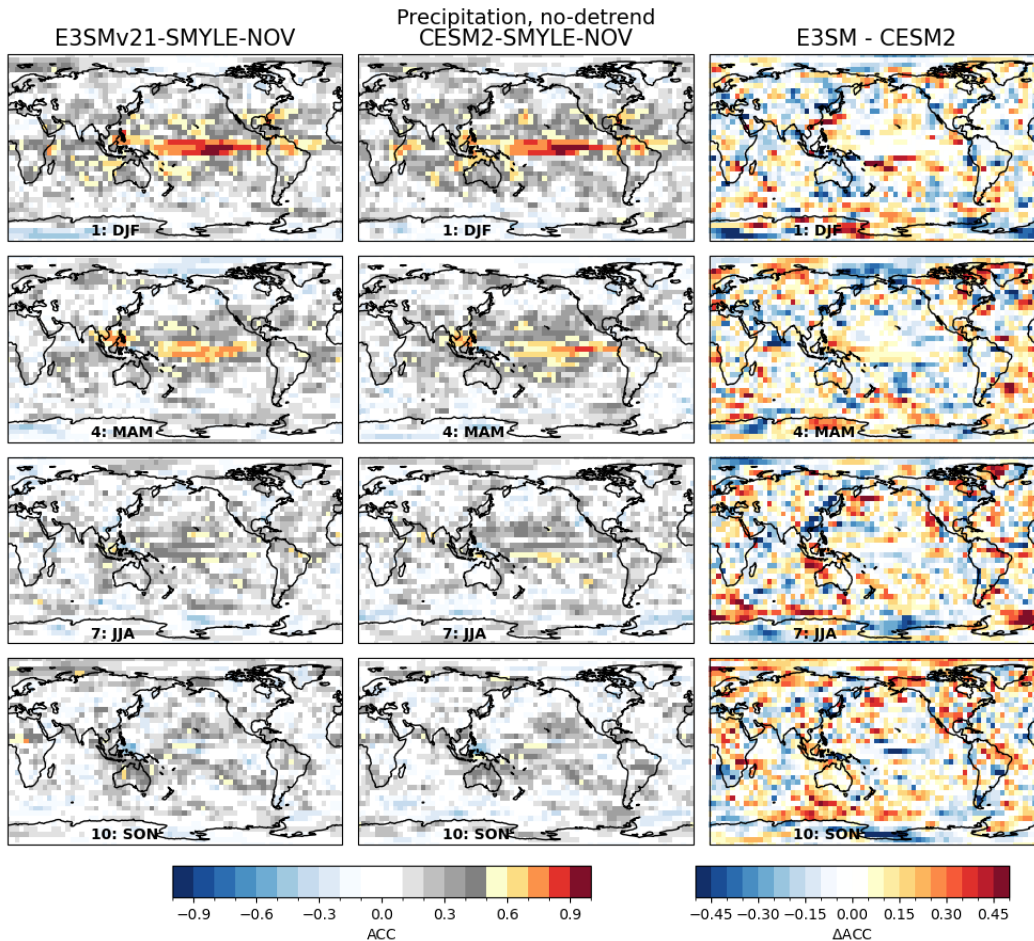
Precipitation, no-detrend
E3SMv21-SMYLE-NOV   CESM2-SMYLE-NOV   E3SM - CESM2

Model/System Sensitivity

- E3SMv2.1-SMYLE system (~10K sim-years)

- Different component models but similar methods as CESM2-SMYLE

- Early analysis suggests no systematic, significant* skill difference for Niño3.4 or hydroclimate over CONUS

May init

Precipitation, no-detrend

E3SMv21-SMYLE-MAY   CESM2-SMYLE-MAY   E3SM - CESM2
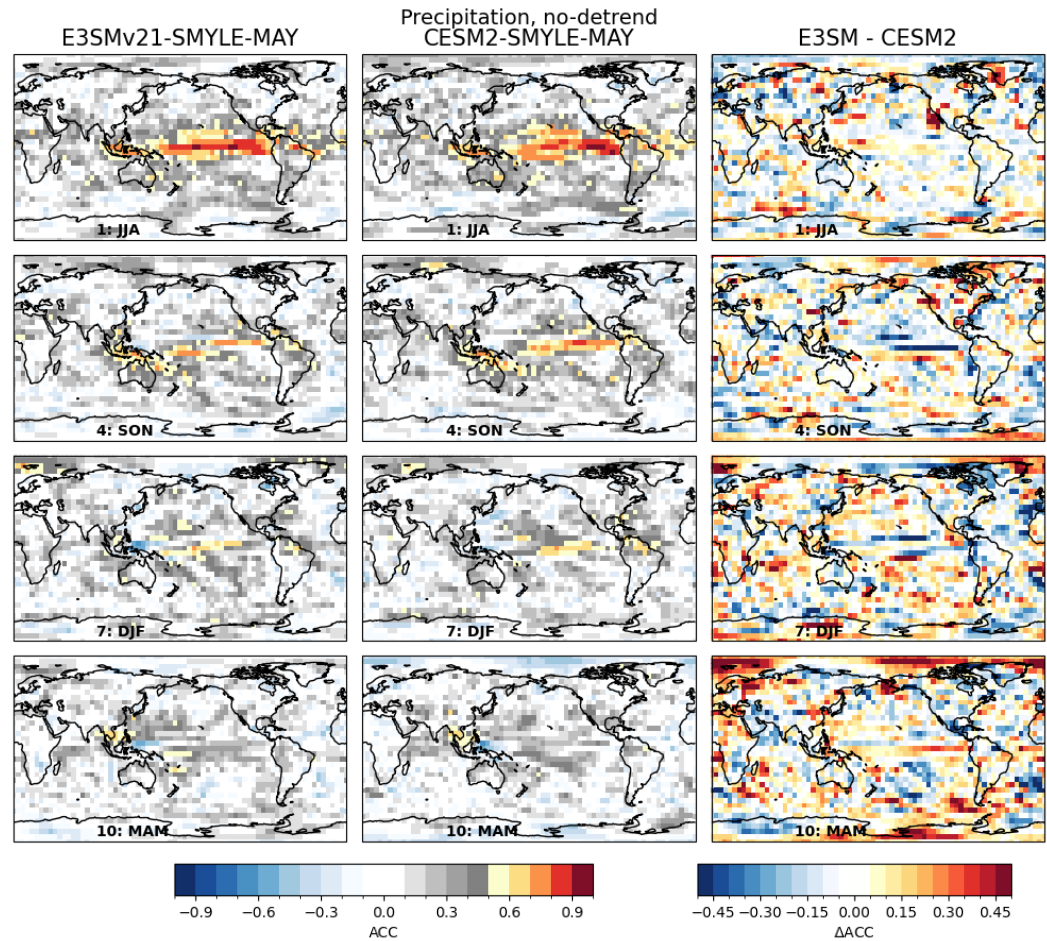
Model/System Sensitivity

- 40-member multi-model system yields only minor improvements (significance uncertain)
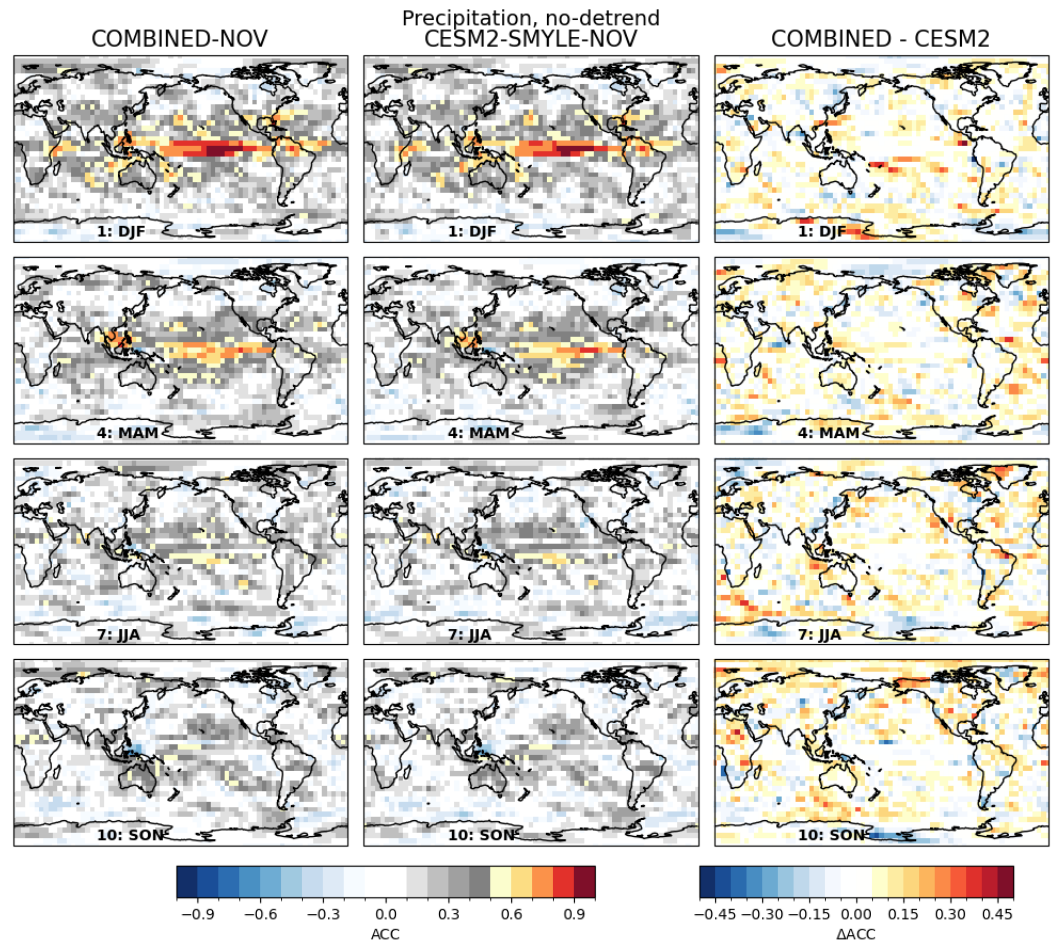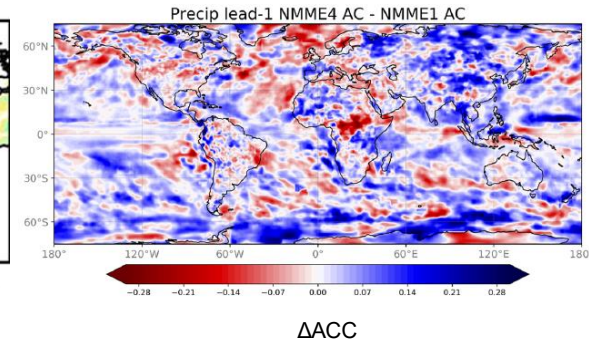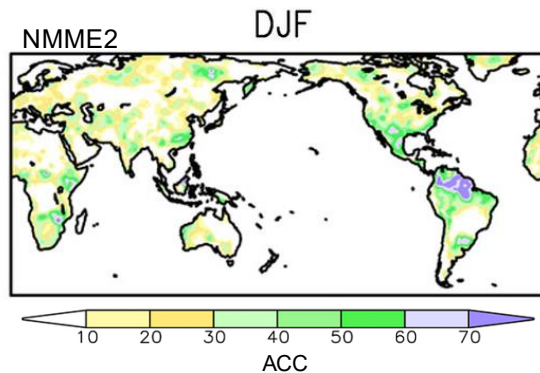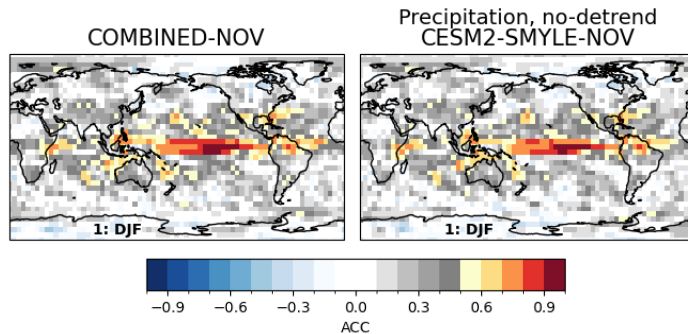

Precipitation, no-detrend

Model/System Sensitivity

- SMYLE skill over CONUS is comparable to NMME

- NMME has shown slight CONUS skill degradation over time (unclear why)

Becker et al. (2014, doi: 10.1175/JCLI-D-13-00597.1,
2020, doi: 10.1029/2020GL087408   )

Model/System Sensitivity

- SMYLE skill over CONUS is comparable to NMME

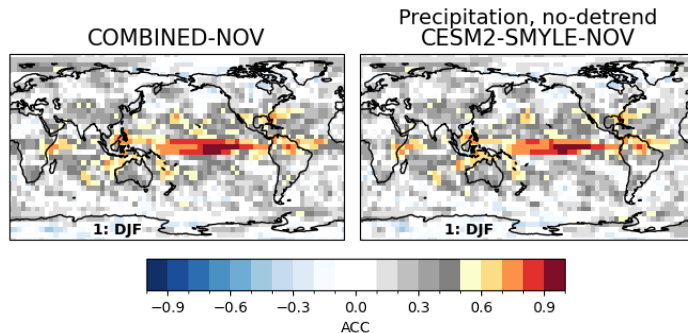- NMME has shown slight CONUS skill degradation over time (unclear why)

How to (systematically) advance understanding of practical predictability limits?
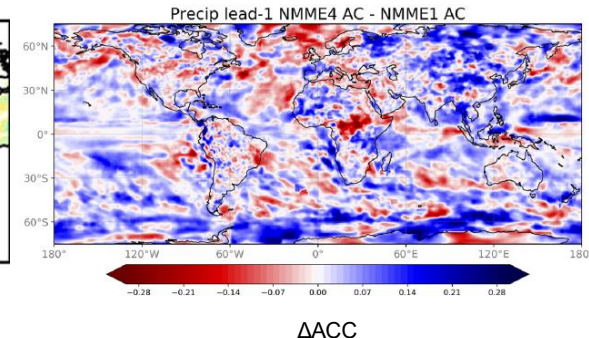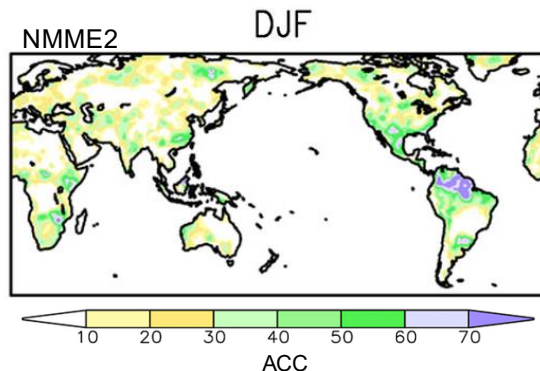
Becker et al. (2014, doi:10.1175/JCLI-D-13-00597.1, 2020, doi:10.1029/2020GL087408   )
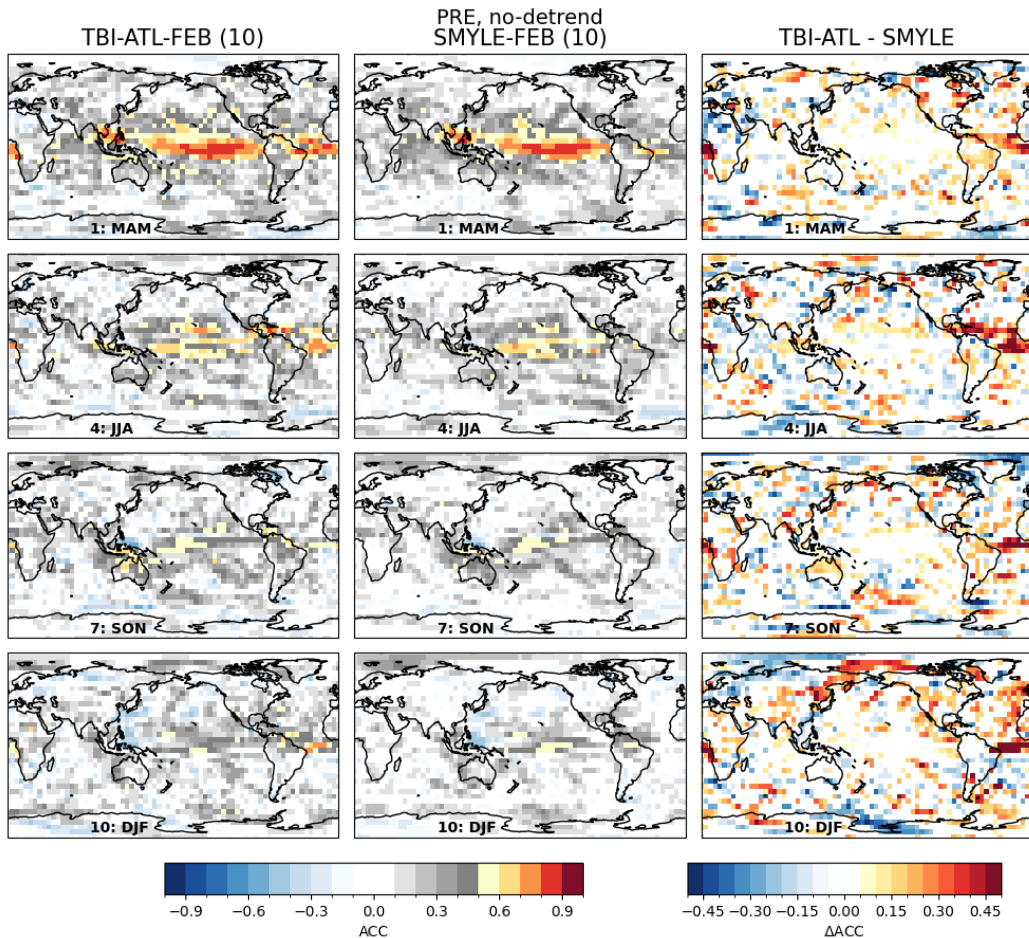
Prediction Pacemakers

- CLIVAR Tropical Basin Interaction MIP (TBIMIP)

Richter et al. (2025, doi:10.5194/gmd-18-2587-2025)

- **TBI-ATL** = CESM2-SMYLE (Feb 1 init) with tropical Atlantic SST nudged to observed anomalies

➔ perhaps low tropical Atlantic predictability contributes to low CONUS hydroclimate skill

  - challenging to interpret
  - expensive

Decadal Hydroclimate Prediction

- CESM2-DP system
  (~15K sim-years)

- New CESM2-DP system shows
  some precipitation skill
  improvement over CESM1-DPLE

  - modified component physics
  - modified ocean initialization
  - *new* land initialization

- To understand skill improvement,
  need to examine prediction system
  developments in isolation (expensive,
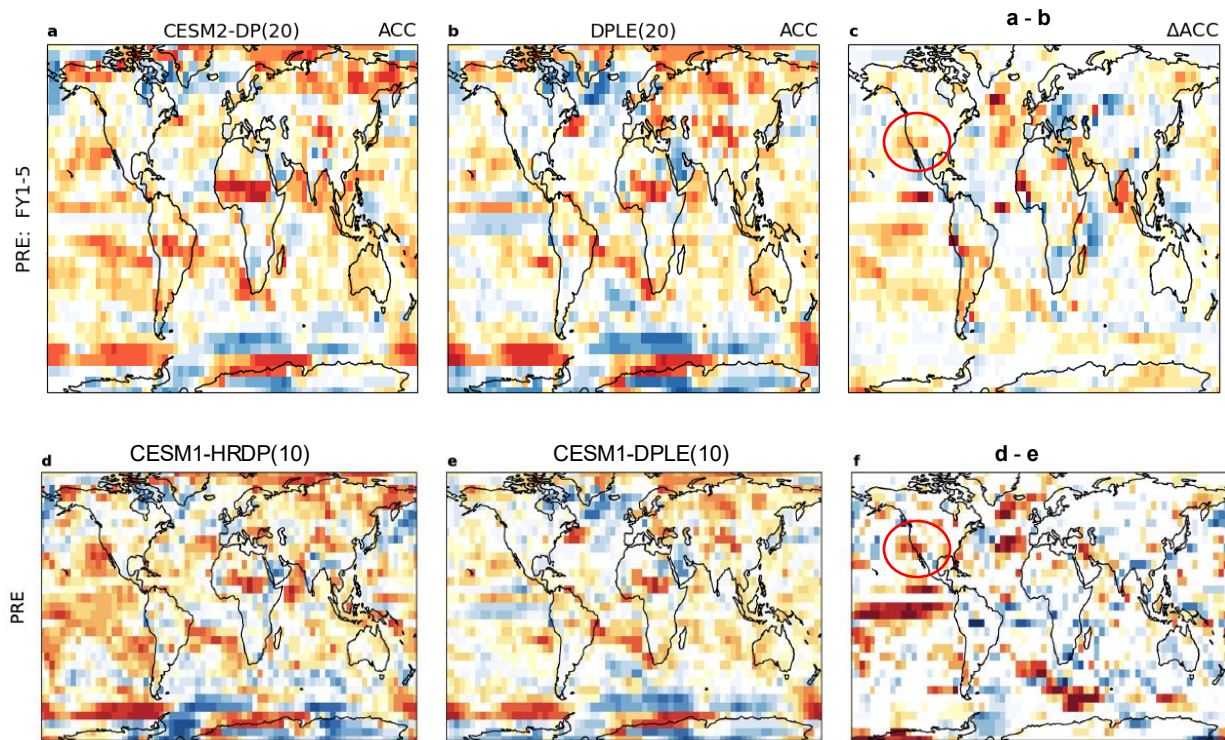 labor-intensive)
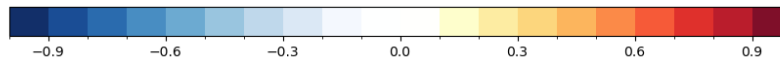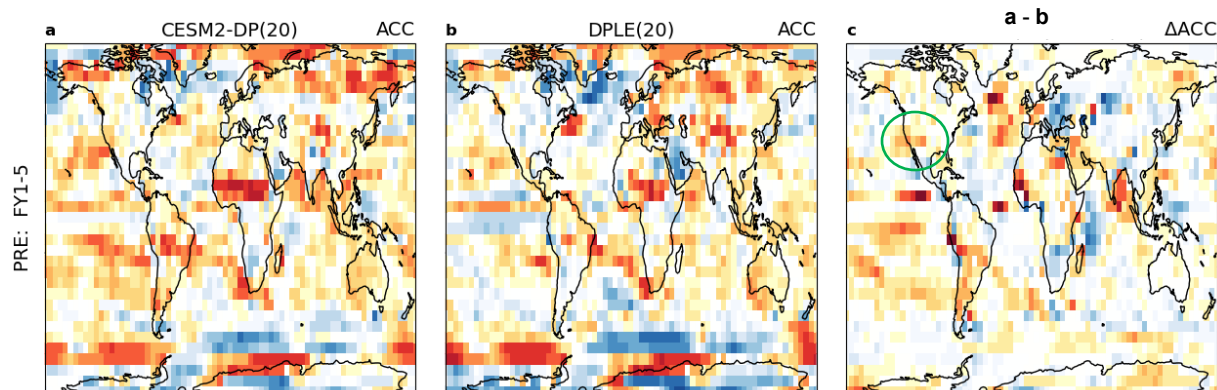
Decadal Hydroclimate Prediction

- CESM2-DP system (~15K sim-years)

- New CESM2-DP system shows some precipitation skill improvement over CESM1-DPLE

  - modified component physics
  - modified ocean initialization
  - *new* land initialization

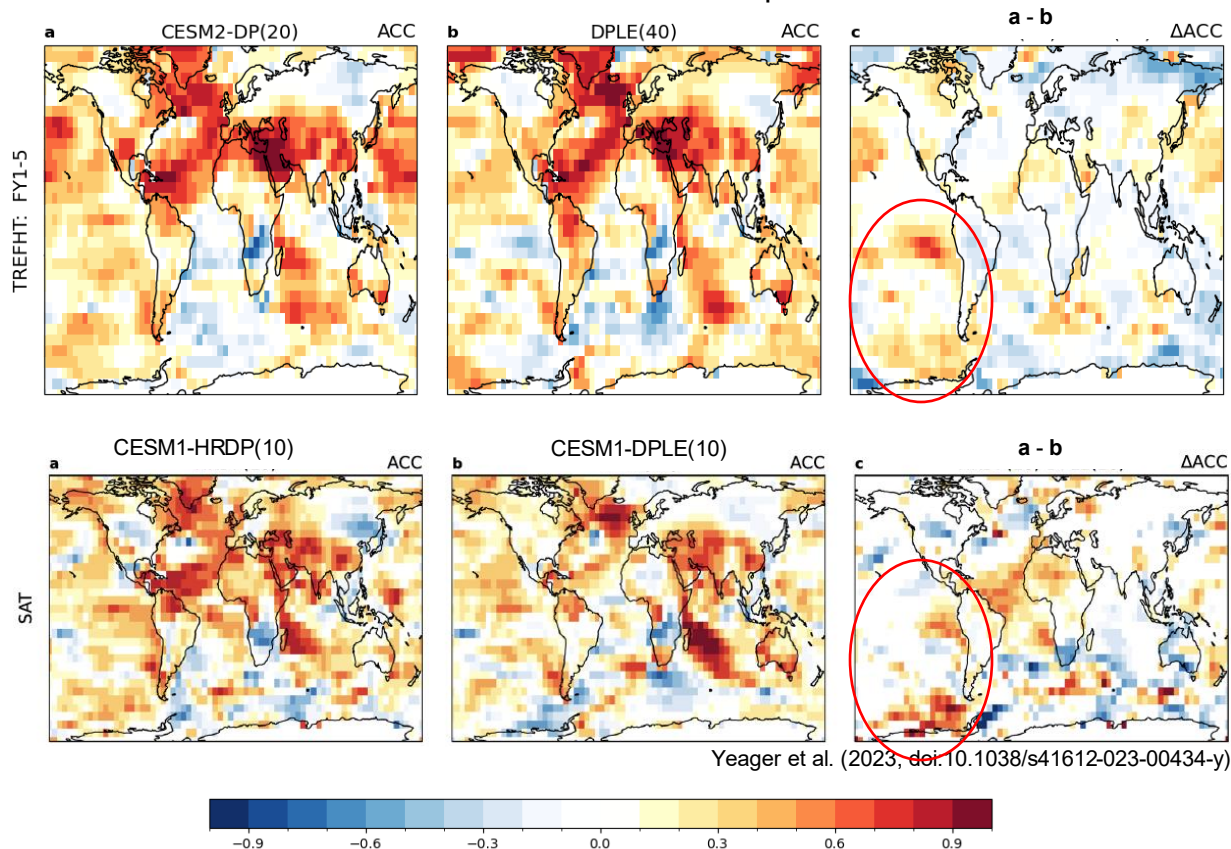- Qualitatively similar to skill improvements seen in CESM1 high-resolution DP system

Yeager et al. (2023, doi:10.1038/s41612-023-00434-y)

Decadal Temperature Prediction

Detrended Surface Temperature Skill

- Similar improvements in SST skill:
  SE tropical Pacific
  Southern Ocean

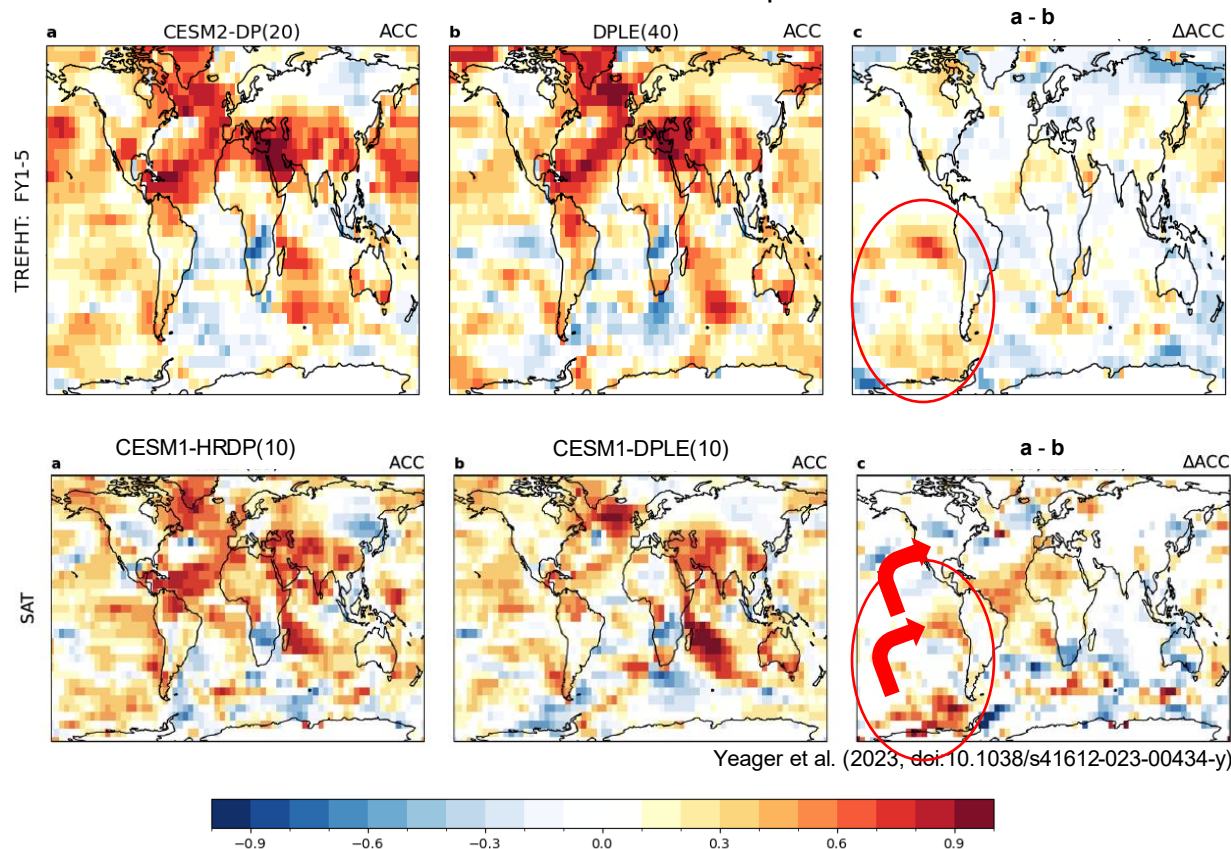Yeager et al. (2023, doi:10.1038/s41612-023-00434-y)

Decadal Temperature Prediction

Detrended Surface Temperature Skill:

- Similar improvements in SST skill:
  SE tropical Pacific
  Southern Ocean

- Hypothesized connection between SO and CONUS hydroclimate via tropical Pacific

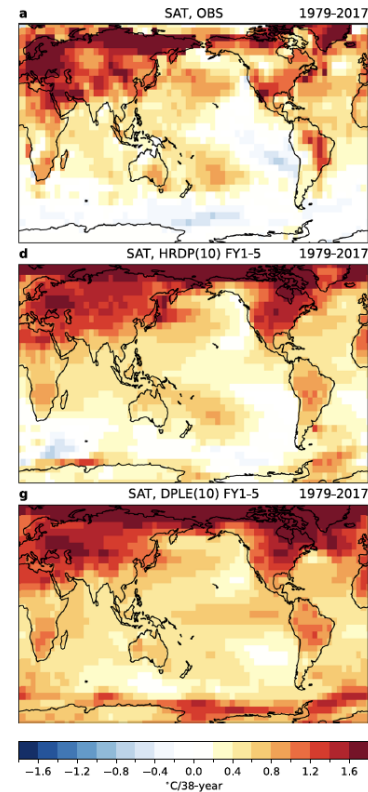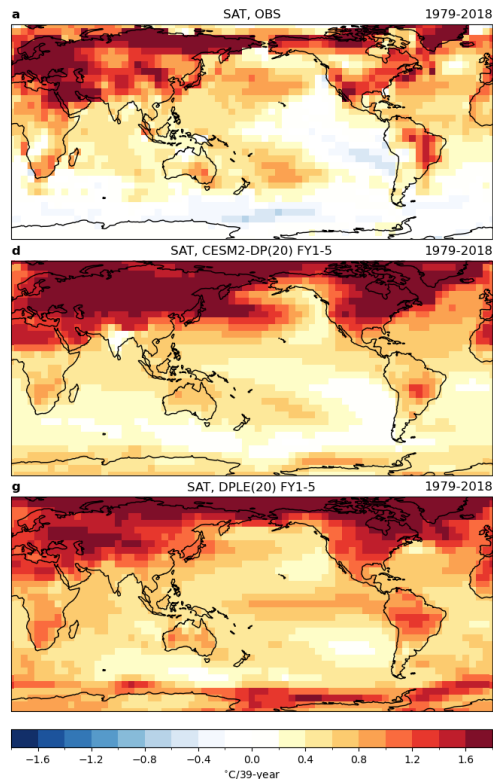Yeager et al. (2023, doi:10.1038/s41612-023-00434-y)

Decadal Temperature Prediction

- SAT skill improvements are related to improved representation of multidecadal trends in the Pacific (and SO)

→ More to this story than just model resolution…

1979-2018 SAT Trends:

Yeager et al. (2023)

Prediction Drift

- Systematic exploration of how predictability limits depend on system design choices (model physics, resolution, parameters, initialization, etc.) is difficult, but **necessary**

- Large reforecast sets generally required to quantify model (biased) lead-dependent climatology before verifying forecast anomalies
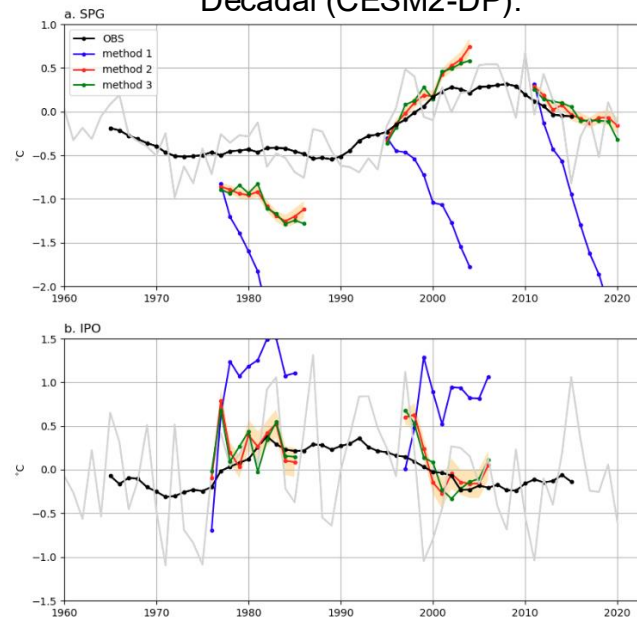
Surface Temperature Bias:
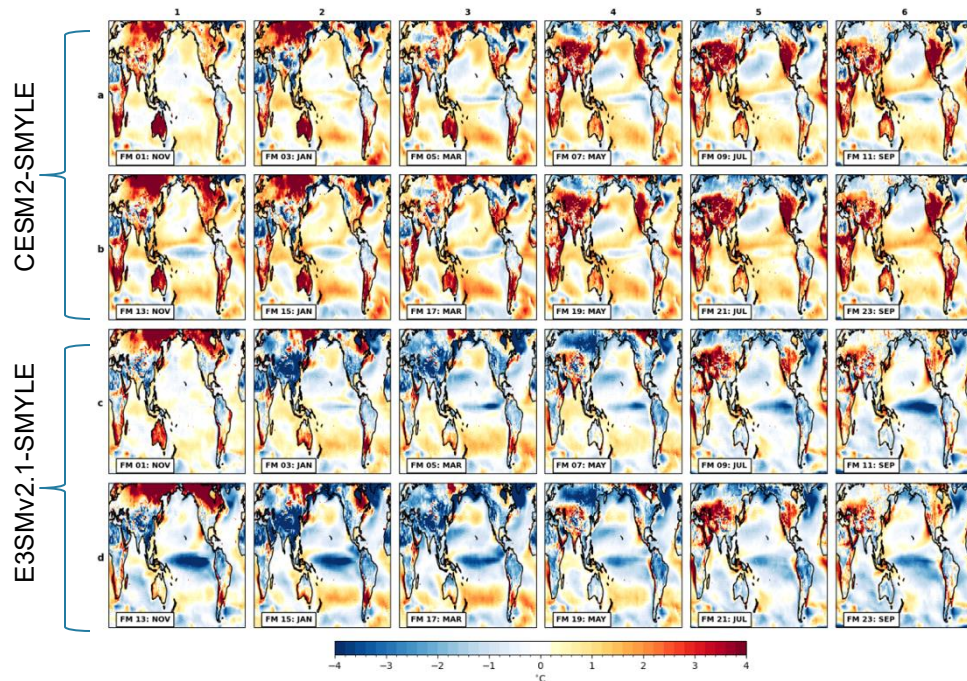
Prediction Drift

- Efficient de-drifting techniques can facilitate expanded case study experimentation with tolerable error

  (Yeager et al., 2025: Efficient Drift Correction of Initialized Earth System Predictions, *submitted*)

Seasonal (CESM2-SMYLE):
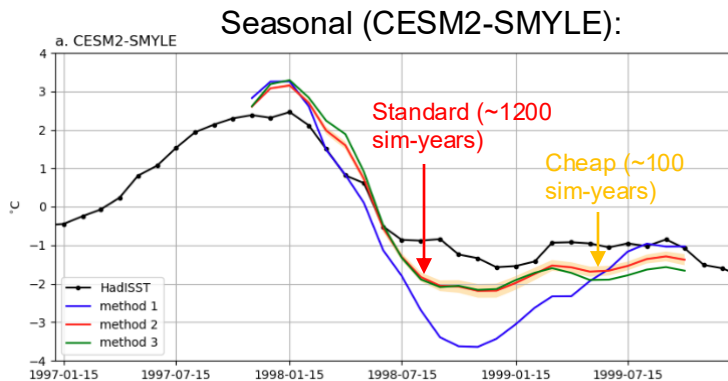
Decadal (CESM2-DP):

Conclusions

- Unclear why high ENSO skill does not yield good skill for seasonal CONUS hydroclimate (inherent limits? system deficiency?)

- To first-order, CESM2 and E3SMv2.1 exhibit similar seasonal predictability (despite very different background mean states); more in-depth, event-focused comparisons may yield useful insight

- Improved decadal prediction skill in CESM2 (relative to CESM1) is promising and appears attributable to SO-related mechanisms revealed by high-res CESM1 efforts (improved trend representation)

- Systematic experimentation is needed to better understand S2D predictability limits
  - coordinated experiments (e.g., CLIVAR TBI, DCPP CMIP7, CESM2/E3SMv2.1)
  - prediction case study experiments to isolate senstivities